Estimation of Dynamic Hedonic Housing Models with Unobserved Heterogeneity*

Susumu Imai textsuperscript1, Modibo Sidibï¿½ textsuperscript2

November 21, 2012

Abstract

In this paper, we estimate a dynamic hedonic model with unobserved housing characteristics. In contrast to the traditional hedonic approach that focuses on the observed housing characteristics, we use the mobility choice from the housing tenure data and the observed hedonic housing characteristics as the main sources of identification. The estimation problem turns out to be much simpler than the location choice literature, where the potential choice set is large. We recover for each housing unit one specific unobserved characteristic as well as individual specific utility shocks. Our approach allows for the unobserved housing characteristics to be correlated with the observed ones. The estimation only requires single market housing data.

We use mobility data on rental apartments from the French Housing Survey. We exploit the strong rent control regulation in France, which makes the rent effectively invariant to changes in local economic conditions. Therefore, we do not need to worry about the endogeneity of rent in the mobility decision due to housing specific unobserved heterogeneity.

* *Preliminary and incomplete. Comments are very welcome.*

[1]Queen's University; imais@econ.queensu.ca

[2]CREST, GATE-LSE, University of Lyon; modibo.sidibe@ensae.fr

Keywords: Hedonic analysis, Identification

JEL Codes: R21, C31, C11.

Introduction

In this paper, we are interested in the identification of individual willingness to pay for housing observed and unobserved characteristics which is central to the understanding of the mechanism of segregation. We propose a new methodology to estimate a housing demand model with unobserved housing characteristics, which also includes unobserved neighborhood attributes.

Most of the literature on location choice, such as **?** or **?**, analyzes individuals housing choices assuming that their choice set includes all housing units in a metropolitan area. In their model, the choice probability of each housing unit is a function of observed and unobserved housing characteristics of all available housing units. Even in the simpler setup of neighborhood choice, the choice probability is a function of observed and unobserved characteristics of all neighborhoods. Estimation of these models is extremely complex due to the large number of choices, since the number of housing units or, even in a simplified model, the number of neighborhoods is usually large, thus causing finite sample difficulties. Often, strong assumptions on the error term of the utility function need to be imposed in order to get reasonable estimates.

The hedonic literature recasts the household location problem into a continuous choice problem of housing or neighborhood attributes. **??** have shown under which set of assumptions a continuous choice model of housing demand can be identified. However, this theoretical literature is based on features that are rare in the real data. As a consequence, there have very few application of this framework.

We bring together the dynamic discrete choice and hedonic literature in a unique framework that allows to identify the structural parameters of both the pricing function and individual preferences. We separate the individual's decision into two choices: the traditional hedonic optimal choice of housing characteristics and the subsequent optimal stopping problem in the decision to stay or move.

We use two features of the data as the main sources of identification. We use the conditional choice probability of whether to stay or move as well as the housing tenure information, along with the observed housing characteristics and the housing location information. The conventional static hedonic approach by **?**, **?**, **?** and others, only uses data on the choice of observed housing characteristics, and thus have difficulties in dealing with unobserved housing characteristics. They have to assume that unobserved housing characteristics are independent of the observed ones. By using both the tenure and mobility choice data and the housing characteristics data, we can allow for correlation between observed and unobserved housing characteristics. We also recover flexibly specified hedonic per period utility function and the individual specific utility shocks as well, when assuming the conventional separability between housing and individual characteristics. Otherwise, the literature requires multimarket data and the assumption that the distributions of observed individual and housing characteristics vary across market while the distribution of unobserved characteristics do not. In our analysis, we only need data on one housing market. The panel dimension of the data will help identification of the unobserved heterogeneties.

The two separate features of the data source are complementary in identifying the model. Mobility and tenure information helps to identify hedonic utility function of housing units because individuals whose deterministic utility of housing consumption is high, stay at the housing unit longer. If we had just used the discrete mobility data, then we would have only been able to identify the finite mixture distribution or parametrized distribution of unobserved heterogeneity of each housing unit. By using additionally housing characteristics data, which comes from the continuous choice of individuals, we are able to recover the unobserved characteristics of each housing unit, and the individual taste shocks as well.

We use data on rental apartments from the French subset of European Community Statistics on Income and Living Conditions (EU-SILC) on the period from 2004 to 2009. We exploit the strong rent control regulation in France, which makes the rent effectively invariant to changes in local economic conditions. Therefore, we do not need to worry about the endogeneity of rent in the mobility decision due to the housing specific unobserved heterogeneity. Therefore, we avoid the need for an instrument for price of a housing consumption, which has always been a difficult issue in the housing literature.

1 Housing demand models with unobserved heterogeneity

1.1 The Literature of Hedonic Model of Housing

We briefly review the literature on housing demand using the hedonic framework. The analysis of hedonic models was pioneered by **?**. In this paper, Rosen suggests a two-step estimation to recover preferences parameters. In the first step, the marginal price function is recovered from the regression of price on attributes. Then, the first-order-conditions (FOC) are used to estimate the utility function. However, as noted by **?**, the second stage suffers from a simultaneity issue. Hence, **?** argue that using a linear approximation of the first order conditions, relying on multi-markets data circumvents the identification issue . In order to achieve identification with multi-market data, it is necessary to assume that the preferences parameters are common across markets while individual heterogeneity varies across markets. Subsequently, **?** and **?** argue that the model is still not identified because unobserved tastes affect both the quantity of an amenity consumed by an individual and its price. Hence, they suggest to use instrumental variables.

The current literature departs from the preceding one in two directions: by using nonparametric methods instead of a linear approximation of the FOC (**??**), and by allowing for unobserved heterogeneity (**?**).

Let's introduce some notation based on **?**'s model to make clear the identification issues. The individual's utility function of a house with observed characteristics z and utility shock ϵ is $U(z, x, c, \epsilon)$ where x is the vector of individual observed characteristics and c is the non-housing consumption. An individual maximizes her utility subject to the following budget constraint:

$$c + P(z) \leq y$$

where the price of non-housing consumption c is normalized to one. Then, the first order condition that allows to recover housing demand is

$$P_z(z) = h(x, z, \epsilon) \equiv \frac{U_z(z, x, y - P(z), \epsilon)}{U_c(z, x, y - P(z), \epsilon)}$$

? show that hedonic models with an additively separable utility function are nonparametrically identified with single market data and present two methods for recovering the structural functions in such models. **?** relax the additivity assumptions, and show that only multimarket data identifies all parameters of the model. As stated before, the underlying assumption for the identification based on multimarket data is that the observed individual or housing characteristics need to be differently distributed across markets but the unobserved individual or housing characteristics need to remain the same. Moreover, another implicit assumption in the models of **?** and **?** is that all housing characteristics z are observable. In this setup, it is important to note that ϵ cannot be interpreted as the unobserved housing characteristic since it is not included in the price function. If the individual knew that the price of unobserved characteristics was zero, she would consume it at the bliss point and thus the distribution of ϵ would be degenerate. If we include ϵ in the price equation, then the price equation becomes

$$P(z, \epsilon) = \delta_z z + \delta_\epsilon \epsilon$$

This is the price equation analyzed in **?**. Applications of this model include **?**, who use non-parametric methods to estimate the coefficients δ_z and interpret the residuals as the unobserved housing characteristics. Once unobserved characteristics are known, then one could just proceed to estimate the parameters in the standard way discussed above. However, to consistently estimate the coefficients of the price equation requires dealing with the correlation of z and ϵ this is way **?**, **?** assume that they are not correlated with each other.

The major difficulty is to allow for correlation between z and ϵ. This requires an instrument, that is, a variable correlated with the observed housing characteristics z but uncorrelated with the unobserved housing characteristics ϵ.

Another strand of the literature, pioneered by **?**, tries to estimate the unobserved housing characteristics directly from the housing choice of the individuals. Denote by z_j and η_j, the observable and unobservable characteristics of a housing unit j. Since η_j are not observable, **?** follow the convention in the literature and specify the housing specific unobserved heterogeneity η_j.

Then, the optimal housing choice of an individual i can be expressed as follows:

$$k = \arg\max_j \{U(z_j, x_i, y_i - P(z_j, \eta_j, w_j), \eta_j, \epsilon_{ij})\}$$

where ϵ_{ij} is the individual i specific utility shock for housing unit j, which is assumed to be i.i.d. extreme value distributed. Now, the price is a function of both the observed characteristic z_j and unobserved characteristics η_j and a price shock w_j. Furthermore, they assume that

$$U(z_j, x_i, p_j, \eta_j) = V(z_j, x_i, p_j) + \eta_j + \epsilon_{ij}$$

where $p_j = P(z_j, \eta_j, w_j)$. Then, household i's choice probability of a housing unit k equals

$$p(i,k) = \frac{\exp[V(z_k, x_i, p_k, \eta_k)]}{\sum_l \exp[V(z_l, x_i, p_l, \eta_l)]}$$

The maximum likelihood estimation is then based on the above household choice probability over all housing units in the data. The estimation follows **?**, and the derivation of the housing unobservable is feasible thanks to the contraction mapping of **?**. Notice that the number of housing specific heterogeneity terms η_j equals the total sample size of housing units, which makes the estimation exercise subject to the finite sample problem.[1]

In a more simpler setting, one could modify the specification of the housing specific unobserved heterogeneity as follows.

$$\eta_j = \xi_n$$

where ξ_n is the neighborhood specific unobserved heterogeneity. Even then, the estimation problem could be subject to the finite sample problem if the number of observations per neighborhood is small, which is typical in a disaggregated neighborhood level data. Furthermore, it is known that the logit choice framework imposes strong functional form (I.I.A.) on the utility function, which may distort the welfare calculation (see **?** for more details).

Another issue in the estimation of the above location model is the endogeneity of the housing characteristics z_k, i.e. their potential correlation with the neighborhood specific unobserved heterogeneity. A proposed solution for the location-related characteristics is to rely on quasi-random variation like geographical boundaries (**??**). However, it is in general difficult to find an appropriate instrument for all other housing characteristics, such as number of rooms, since one

[1] **?**, after implementing the above estimation algorithm, hint that a potential approach that is not subject to the finite sample bias would be to assume that η_j follows a distribution that is nonparametrically estimated, for instance finite mixture. This would certainly avoid the finite sample problem mentioned above, but would also create an additional issue that $V(z_j, x_i, p_j) + \eta_j$ can no longer be obtained *à la* BLP style in the first stage using the contraction mapping, which may add complexity in the estimation.

has to find an instrument that is correlated with the observed characteristics but uncorrelated with the unobserved characteristics.[2]

The literature that is the closest to our paper is **?**. This approach is based on the sorting model of **?** which predicts the positive correlation between household's neighborhood quality and average permanent income of household's residence. The issue there is that the permanent income is not observed, and additional instruments are required since **?** predict a negative within neighborhood correlation between observed and unobserved local income. **?** estimates a dynamic model of location choice with moving cost. The estimation proceeds in two steps. In the first, a dynamic model of migration is estimated and a location fixed effect is recovered. In the second, the location fixed effect is decomposed between observed and unobserved determinants using the same argument as **?**. Nonetheless, in presence of individual sorting, the second step may still suffer from an endogeneity bias.

1.2 Our Dynamic Model

Instead of identifying the structural parameters solely from the residential choice of the individuals, we additionally use the dynamic mobility choice, i.e. whether to stay in the same residential unit or move.[3] After a mobility, we model the subsequent tenure choice of individuals. Consider a set of private tenants. Let τ be the length of stay in the neighborhood, and Z_i and w_i be the vector of the observed and unobserved characteristics of a housing unit i. We assume that the relative rent of apartment i depends on the length of stay τ, housing observed attributes and an unobserved component w_i :

$$r(\tau, Z_i, w_i)$$

[2] For example, **?** uses prices of far away neighborhoods as instruments, where the implicit assumption is that the unobserved characteristics of the far away neighborhoods are not correlated with prices.

[3] A list of notation is available in the appendix A.

In our framework, we consider the relative rent as an intuitive way to introduce the individual choice problem. That is, the rent considered here summarizes the fact that, when an individual stays longer, his rent relative to the cost of other housing units with lower length of stay decreases.

Furthermore, we assume that the relative rent of a housing unit is determined as follows

$$r(\tau, Z_i, w_i) = r_0(Z_i, w_i)g(\tau)$$

where r_0 is the rent at initial period, and $g(\tau)$ represents the over time change in relative rent, which is entirely determined by the regulated uniform growth rate of rent, where t_0 is the initial period. We let the period zero rent equation to be linearly separable between observed and unobserved heterogeneity, i.e.

$$r_0(Z, w) = h(Z) + v(w)$$

Where $h(Z)$ is a function of observed attributes, and $v(w)$ is the unobserved part of the rent equation. Then, the individual in each period makes the choice between staying and moving out of the house. Let the per period utility of staying in a house be specified as.

$$\gamma_1 r_i(\tau) + \frac{1}{2}\gamma_2 r_i(\tau)^2 + \zeta_{i\tau} + \epsilon_{0\tau i}$$

where the first two terms represent the negative utility from paying the rent r_i, and the fourth term represents the per period utility shock. The third term is a function of observable and unobservable characteristics:

$$\zeta_{i\tau} = [b(Z_i, X_{i,\tau}, u_{Z,i}) + u(w_i)]c(X_{i,\tau})u_i$$

where $c()$ is a function of individual observed characteristics $X_{i,\tau}$, $b()$ is the utility component for observed attributes, which is a function of housing observed characteristics Z_i and individual observed characteristics $X_{i,\tau}$. $u_{Z,i}$ and $u(w_i)$ are utility shocks that captures the fact that the effects of the observed Z, and unobserved w_i differ per individuals. u_i is the shock for the marginal utility of housing characteristics.

As we discussed before, we do not observe each component of the vector w_i. Given parameters of the rent equation $h()$, we can only recover $v(w_i)$, thus approximate the utility component of unobservable characteristics as follows:

$$u(w_i) \approx u_{v,i} v(w_i)$$

That is, the utility components of the renting and the preference function are assumed to be proportional. Thus, we estimate $v(w_i)$ the effect of unobserved heterogeneity on the price, and the proportionality parameter $u_{v,i}$ between $u(w_i)$ and $v(w_i)$. As a consequence, we can drop the term w_i without loss of generality. This setup is similar to the one where the utility function is linearly separable in other goods $x = y - r_i$ and a function of housing characteristics Z. [4]

We denote the tenure invariant component of the rent as $r_i(0)$. Since the rent of all subsequent periods is a deterministic function of the initial rent, all the subsequent mobility choices are a function of $r_i(0)$. The choice specific value of staying is

$$\begin{aligned} V_{s,\tau}(r_i(0), X_{i\tau}, Z_i, \zeta_i(v_i, u_Z, u_i, u_v), \epsilon_{s\tau i}) &= \gamma_1 r(\tau, Z_i, v_i) + \frac{1}{2}\gamma_2 r(\tau, Z_i, v_i)^2 + \zeta_{i0} + \epsilon_{s\tau i} \\ &\quad + \beta E V_{\tau+1}(r_i(0), X_{i\tau+1}, \zeta_i, Z_i, \epsilon_{s(\tau+1)i}) \end{aligned}$$

[4]For notation convenience, in the rest of the paper, we drop the subscript i in the notation of $u_{Z,i}$, $u_{v,i}$.

The value of moving out and rent another privately owned housing unit is

$$V^R_{m\tau}(X_{i\tau}, u_Z, u_i, u_v, \epsilon_{m\tau i}) = \bar{V}^R(X_{i\tau}, u_Z, u_i, u_v) + \epsilon_{R\tau i}$$

where $\bar{V}_R$ is the deterministic value function of moving out that will be defined later. β is the discount factor. One can now define the value function associated to a mobility and the decision to buy a house or move to public housing.

We model homeownership as a purely reduced form since we can not overcome the endogeneity of prices. The value of homeownership is defined as

$$V^h_{m\tau}(X_{i\tau}, u_Z, u_i, u_v, \epsilon_{h\tau i}) = \bar{V}^h_m(X_{i\tau}, u_Z, u_i, u_v) + \epsilon_{h\tau i}$$

Similarly, the value of moving into public housing is written as:

$$V^P_{m\tau}(X_{i\tau}, u_Z, u_i, u_v, \epsilon_{h\tau i}) = \bar{V}^P_m(X_{i\tau}, u_Z, u_i, u_v) + \epsilon_{p\tau i}$$

As a consequence, the value of moving out can be defined as:

$$V_{m\tau}(X_{i\tau}, u_Z, u_i, u_v, \epsilon_{m\tau i}) = \max\left[V^R_{m\tau}(X_{i\tau}, u_Z, u_i, u_v, \epsilon_{R\tau i}), V^H_{m\tau}(X_{i\tau}, u_Z, u_i, u_v, \epsilon_{h\tau i}), V^P_{m\tau}(X_{i\tau}, u_Z, u_i, u_v, \epsilon_{h\tau i})\right] \tag{1}$$

Let's rewrite $\epsilon_{m\tau i}$ as the error term associated to a mobility, and assume that both $\epsilon_{s\tau i}$ and $\epsilon_{m\tau i}$ are assumed to be i.i.d. extreme value distributed. As a consequence:

$$\begin{aligned} EV_\tau(r_i(0), X_{i\tau}, Z_i, v_i, u_Z, u_i, u_v) &= \log\left\{\exp\left[\bar{V}_{s\tau}(r_i(0), X_{i\tau}, Z_i, v_i, u_Z, u_i, u_v)\right]\right. \\ &+ \left.\exp\left[\bar{V}_{m\tau}(X_{i\tau}, u_Z, u_i, u_v)\right]\right\} \end{aligned}$$

Where $\bar{V}_{s\tau}$ denotes the deterministic value of staying:

$$\bar{V}_{s\tau}(r_i(), Z_i, \tau, X_{i\tau}, v_i) = \gamma_1 r(\tau, Z_i, w_i) + \frac{1}{2}\gamma_2[r(\tau, Z_i, w_i)]^2 + \zeta_{i\tau} + \beta EV_{\tau+1}(r_i(0), \tau+1, X_{i\tau+1}, \zeta_i).$$

Then, the probability of leaving the house at period τ is

$$p_\tau(r_i(0), Z_i, X_{i\tau}, \zeta_i) = \frac{\exp\left[\bar{V}_m(X_{i\tau}, Z_i, \zeta_i)\right]}{\exp[\bar{V}_{s\tau}(r_i(0), \zeta_{i0}, \tau, X_{i\tau}, Z_i)] + \exp[\bar{V}_m(X_{i\tau}, Z_i, \zeta_i)]} \tag{2}$$

At the initial period, 0 when an individual with the characteristics X_{i0} is looking for a house, she tries to find a housing unit that maximizes the following value function with respect to observed housing characteristics Z_i and unobserved housing characteristics v_i.

$$EV_0(X_{i0}, u_i, u_Z, u_v) = \arg\max_{\{Z_i, v_i\}} EV_0(X_{i0}, u_Z, v_i, u_i, r_i(0)) \tag{3}$$

As a consequence, the deterministic value of moving out can be derived as:

$$\bar{V}_{m\tau}(X_{i\tau}, u_Z, u_i, u_v) = EV_0(X_{i\tau}, u_Z, u_i, u_v) - MC(X_{i\tau})$$

where $MC(X_{i\tau})$ is the mobility cost for an individual with characteristics X_i at period τ.

Here, we impose a restriction that there is only one unobserved characteristic v_i. That is, if there are several unobserved characteristics of the rental unit, individuals do not choose optimally the quantity of each of them, and thus they can be considered as an index.

2 Data

We draw our data from the French subset of European Community Statistics on Income and Living Conditions (EU-SILC) on the period from 2004 to 2009.

EU-SILC was launched in 2004 with the aim to follow changes in individual characteristics over time along dimensions like income, and living conditions. As a consequence, it provides a longitudinal data that allows to track the status of individuals in the labor market (income, employment status), housing market (tenure, characteristics of housing units, rent), and geographical mobility over a maximum period of 9 years. In addition, it provides, the year of arrival of the household in the current housing unit, allowing to study

Unfortunately, the data has several drawbacks that deserve a mention. First, it is a rotating panel with the aim to replace a ninth of the sample every period, and individuals who are lost over time. As a consequence, we are not able to follow a majority of individuals in the initial sample can not be followed over a long period of time. Second, the survey is not mandatory after 4 years of observations, with yields a high degree of attrition the fifth year of the survey. Finally, there is a great deal of attrition.

[INCOMPLETE]

2.1 Transitions in the Housing Market

2.2 Descriptive Statistics

3 Identification

3.1 The French rental market

The French rental market is characterized by a large number of regulations. As noted by **?**, these regulations are biased towards tenants. The major regulation takes the form of a winter recess during which landlords have to cope with unpaid bills. The second major characteristic, which is more of interest for our work, is related to the presence of a rent evolution regulation. Landlords can freely set the initial rent of their dwelling. After the initial period, the rent evolution is given by a rent index calculated by the French Statistical Institute (INSEE) by using the construction cost index.[5] This regulation has mainly two consequences. First, it reduces significantly the degree of uncertainty on future rents, and hence operates as an insurance against shocks of rent. A potential implication is to decrease the value of homeownership compared to renting. The homeownership rate in 2006 in France, close to 56%, is lower than in most of the developed countries.[6] Second, the rent regulation creates an exogenous variation paths for relative rent that is independent of market conditions. This is very useful in order to circumvent the endogeneity of rent.

3.2 Identification of the model

The identification is based on the optimal housing choice of an individual. That is, at the beginning of the period, an individual chooses a housing unit which maximizes he expected utility. Let's denote $v_i = v(w_i)$. Given the data on rents and housing observed characteristics, we can recover v_i from the initial rent equation as follows.

$$v_i = r_i(0) - h(Z_i)$$

We then use the following F.O.C. of the optimal choice of housing characteristics. The sensitivity of housing demand to the housing unobserved v_i, and observed Z_i characteristics can be written respectively:

[5]A change was introduced in 2006, the rent index is now based on the consumer price index for all goods except tobacco and shelter.

[6]The exceptions are very specific housing markets: Germany (price fallout after reunification boom), Netherlands (extremely high proportion of public housing).

$$\frac{\partial EV(r_i(0), Z_i, v_i, u_{Z,i}, u_{v,i}, u_i, X_i, \theta)}{\partial v_i} = -\frac{\partial EV(r_i(0), Z_i, v_i, u_{Z,i}, u_{v,i}, u_i, X_i, \theta)}{\partial r_i(0)} \frac{\partial r_i(0)}{\partial v_i}$$

$$\frac{\partial EV(r_i(0), Z_i, v_i, u_{Z,i}, u_{v,i}, u_i, X_i, \theta)}{\partial Z_i} = -\frac{\partial EV(r_i(0), Z_i, v_i, u_{Z,i}, u_{v,i}, u_i, X_i, \theta)}{\partial r_i(0)} \frac{\partial r_i(0)}{\partial Z_i}$$

Now, notice that the same can be derived using the utility shock to marginal utility

$$\begin{aligned}
\frac{\partial EV(r_i(0), Z_i, v_i, u_{Z,i}, u_{v,i}, u_i, X_i, \theta)}{\partial \mu_{i\tau}} &= \beta^\tau \frac{\partial EV(r_i(\tau), Z_i, v_i, u_{Z,i}, u_{v,i}, u_i, X_{i\tau}, \theta)}{\partial \mu_{i\tau}} \\
&\times \; P(s \geq \tau, r_i, Z_i, v_i, u_{Z,i}, u_{v,i}, u_i, X_{i,1:\tau}) \\
&= \beta^\tau \frac{exp(\overline{V}_{0\tau})}{exp(\overline{V}_{0\tau}) + exp(\overline{V}_{1\tau})} \times P(s \geq \tau, r_i, Z_i, v_i, u_{Z,i}, u_{v,i}, u_i, X_{i,1:\tau}) \\
&= \beta^\tau P(s > \tau, r_i, Z_i, v_i, u_{Z,i}, u_{v,i}, u_i, X_{i,1:\tau})
\end{aligned}$$

Where s is an index, and P is the survival probability that depends on $X_{it_0^\tau}$, the sequence of individual characteristics from the initial period to the period τ.

Hence, considering the relationship between ζ and the unobserved housing characteristics v_i

$$\begin{aligned}
&\frac{\partial EV(r_i(0), Z_i, v_i, u_{Z,i}, u_{v,i}, u_i, X_i, \theta)}{\partial v_i} \\
= &\sum_{\tau=1}^{T} \beta^\tau P(s > \tau, r_i, Z_i, v_i, u_{Z,i}, u_{v,i}, u_i, X_{i,1:\tau}) c(X_{i,\tau}) \times u_{v,i} u_{ci}
\end{aligned}$$

Furthermore, notice from the same relationship as before

$$
\begin{aligned}
& \frac{\partial EV(r_i(0), Z_i, v_i, u_{Z,i}, u_{v,i}, u_i, X_i, \theta)}{\partial Z_i} \\
= & \sum_{\tau=1}^{T} [\beta^\tau P(s > \tau, r_i, Z_i, v_i, u_{Z,i}, u_{v,i}, u_i, X_{i,1:\tau}) \times b_Z(Z_i, X_{i,\tau}, u_{Z,i})] c(X_{i,\tau}) u_{ci}
\end{aligned}
$$

Therefore, assume the set of observed characteristics is of size J. We obtain for each element j of vector Z

$$
\frac{\sum_{\tau=1}^{T} [\beta^\tau P(s > \tau, r_i, Z_i, v_i, u_Z, u_v, u_i, X_{i,1:\tau}, Z_i, X_{i\tau}, u_Z) c(X_{i,\tau})]}{\sum_{\tau=1}^{T} \beta^\tau P(s > \tau, r_i, Z_i, v_i, u_Z, u_v, u_i, X_{i,1:\tau})} \tag{4}
$$

Notice that there are $J+1$ unknown variables, $u_{v,i}$, $u_{Z,i}$ with only J equations. Hence, without loss of generality, we normalize u_{z0} to be zero. Then, $u_{v,i}$ can be derived [7] Then, u_v can be derived

$$
u_v = \frac{b_{Z_j}(Z, X_1, u_z) c(X_1) + \sum_{\tau=2}^{T} \beta^{\tau-1} \int_{X_{1:\tau}} p\left(s > \tau, r\left(0\right), Z, X_{1:\tau}\right) b_{Z_j}(Z, X_\tau, u_z) c(X_\tau) dF\left(X_{1:\tau}|X_1\right)}{c(X_1) + \sum_{\tau=2}^{T} \beta^{\tau-1} \int_{X_{1:\tau}} p\left(r\left(0\right), Z, X_{1:\tau}\right) c(X_\tau) dF\left(X_{1:\tau}|X_1\right)} \tag{5}
$$

Then, given Z_i, $u_{z,i}$ is determined that this holds for any Z_j. That is, $u_{z,i}$ satisfies

$$
\begin{aligned}
& const \\
= & b_{Z_j}(Z, X_1, u_z) c(X_1) + \sum_{\tau=2}^{T} \beta^{\tau-1} \int_{X_{1:\tau}} p\left(s > \tau, r\left(0\right), Z, X_{1:\tau}\right) b_{Z_j}(Z, X_\tau, u_z) c(X_\tau) dF\left(X_{1:\tau}|X_1\right)
\end{aligned}
$$

Lastly, we need to derive the utility shock parameter u_{ci}.

[7] For notation convenience, we ignore the arguments of the moving out probability.

$$u_c = \frac{h_{Z_j}(Z)\left[\gamma_1 + \gamma_2 r(1) + \sum_{\tau=2}^{T} \beta^{\tau-1} \int_{X_{1:\tau}} p\left(s > \tau, r\left(0\right), Z, X_{1:\tau}\right) dF\left(X_{1:\tau}|X_1\right)\left[\gamma_1 + \gamma_2 r(0) g(\tau)\right] g(\tau)\right]}{b_{Z_j}(Z, X_1, u_z) c(X_1) + \sum_{\tau=2}^{T} \beta^{\tau-1} \int_{X_{1:\tau}} p\left(s > \tau, r\left(0\right), Z, X_{1:\tau}\right) b_{Z_j}(Z, X_\tau, u_z) c(X_\tau) dF\left(X_{1:\tau}|X_1\right)} \tag{6}$$

$$u_v u_c = \frac{h_{Z_j}(Z)\left[\gamma_1 + \gamma_2 r(1) + \sum_{\tau=2}^{T} \beta^{\tau-1} \int_{X_{1:\tau}} p\left(s > \tau, r\left(0\right), Z, X_1, X_\tau\right) dF\left(X_\tau|X_1\right)\left[\gamma_1 + \gamma_2 r(0) g(\tau)\right] g(\tau)\right]}{c(X_1) + \sum_{\tau=2}^{T} \beta^{\tau-1} \int_{X_{1:\tau}} p\left(s > \tau, r\left(0\right), Z, X_1, X_\tau\right) c(X_\tau) dF\left(X_\tau|X_1\right)} \tag{7}$$

Equations 5 and 6 determine the unobserved heterogeneity, given the parameters of the model. This is how the unobserved heterogeneity is recovered from the optimal choice of the housing characteristics, observed and unobserved, which are similar to the traditional hedonic models. The above specification implies that the unobserved heterogeneity is a following function

$$u = u\left(Z, r(0), X_1\right)$$

That is, given the observed characteristics of the individual, the unobserved characteristics are uniquely determined so as to make him choose the housing characteristics, observed and unobserved, that we see in the data or recover from the data and the model. However, u does not uniquely determine $(Z, r(0), X_1)$. That is, individuals who have the same unobserved characteristics could choose different housing characteristics because of the difference in the observed characteristics at the time of choice. Assuming that X_τ has an AR specification,

$$p\left(\tau, r\left(0\right), Z, X_\tau, u\left(Z, r(0), X_1\right)\right) = \frac{exp(\bar{V}_0(\tau, r(0), Z, X_\tau, u(Z, r(0), X_1)))}{exp\left(\bar{V}_0()\right) + exp(\bar{V}_1(X_\tau, u\left(Z, r(0), X_1\right)))} \tag{8}$$

which is the mobility equation, i.e.

$$p\left(\tau, r\left(0\right), Z, X_{\tau}, u\left(Z, r(0), X_1\right)\right)$$

is the conditional hazard. Because of this,

$$\bar{V}_{0i}(r_i(\tau), Z_i, X_{i,1:\tau}, u, \tau) - \bar{V}_1(X_{i,1:\tau}, u) = \log p\left(r_i, Z_i, X_{i,1:\tau}\right) - \log\left(1 - p\left(r_i, Z_i, X_{i,1:\tau}\right)\right), \tau)$$

Furthermore,

$$EV(r_i(\tau), Z_i, X_{i,1:\tau}, u) = \bar{V}_1(X_{i,1:\tau}, u) - \log p\left(r_i, Z_i, X_{i,1:\tau}\right)$$

Therefore, if we let the functional form specification of $\bar{V}_1$ to be as follows:

$$\bar{V}_1(X_{1:\tau}, u) = \eta(X_{\tau}, u),$$

where

$$\begin{aligned} \eta(X_{\tau}, u) &= Max\left\{EV(r_i(\tau), Z_i^*, X_{i,1:\tau}, u) + v_1, EV_{nr}\left(X_{i,1:\tau}, u\right) + v_2\right\} \\ \eta(X_{\tau}, u) &= \log\left[\exp\left(EV(r^*, Z_i^*, X_{\tau}, u)\right) + \exp\left(EV_{nr}\left(X_{i,1:\tau}, u\right)\right)\right] \end{aligned}$$

Because

$$p_r\left(X_{\tau}, u\right) = \frac{\exp\left(EV(r^*, Z_i^*, X_{\tau}, u)\right)}{\exp\left(EV(r^*, Z_i^*, X_{\tau}, u)\right) + \exp\left(EV_{nr}\left(X_{i,1:\tau}, u\right)\right)}$$

$$\eta(X_{\tau}, u) = EV(r^*, Z_i^*, X_{\tau}, u) - \log p_r\left(X_{\tau}, u\right)$$

where

$$\log p_r\left(X_{\tau}, u\left(X_1, r, Z\right)\right) = \log p_r\left(X_1, r, Z, X_{\tau}\right)$$

then, for any $\tau > 1$,

$$\begin{aligned} & \left[b\left(Z, X_\tau, u_Z\right)+u_v\left(r_i(0)-h(Z)\right)\right] c(X_\tau) u_c+\gamma r_0 g(\tau)-\eta\left(X_\tau, u\right) \\ & +\beta \int_{X_{\tau+1}}\left[\eta\left(X_{\tau+1}, u\right)-\log \left(1-p\left(r, Z, X_1, X_{\tau+1}\right)\right)\right] d F\left(X_{\tau+1} \mid X_\tau\right) \\ = & \log \left(p\left(\tau, r\left(0\right), Z, X_1, X_\tau\right)\right)-\log \left(1-p\left(\tau, r\left(0\right), Z, X_1, X_\tau\right)\right) \end{aligned}$$

Then, changes in the RHS over tenure τ identifies the coefficient γ. To see why, consider the case where $X_\tau = X_1$. Then,

$$\begin{aligned} & \left[b\left(Z, X_1, u_Z\right)+u_v\left(r-h(Z)\right)\right] c(X_1) u_c+\gamma r_0 g(\tau)-\eta\left(X_1, u\right) \\ & +\beta \int_{X_{\tau+1}}\left[\eta\left(X_{\tau+1}, u\right)-\log \left(1-p\left(r, Z, X_1, X_{\tau+1}\right)\right)\right] d F\left(X_{\tau+1} \mid X_1\right) \\ = & \log \left(p\left(\tau, r, Z, X_1, X_1\right)\right)-\log \left(1-p\left(\tau, r, Z, X_1, X_1\right)\right) \end{aligned}$$

Notice that because $(Z, r - h(Z))$ is the optimal choice given the observed characteristic X_1 and unobserved characteristic u

$$\begin{aligned} \eta\left(X_1, u\right) = & \left[b\left(Z, X_1, u_Z\right)+u_v\left(r-h(Z)\right)\right] c(X_1) u_c+\gamma r_0 g(\tau) \\ & +\beta \int_{X_{\tau+1}}\left[\eta\left(X_{\tau+1}, u\right)-\log \left(1-p\left(2, r, Z, X_1, X_{\tau+1}\right)\right)\right] d F\left(X_{\tau+1} \mid X_1\right) \\ & -\log p_r\left(X_\tau, u\right) \end{aligned}$$

Here, we assume that when moving into a new apartment, the individual has to stay there at least for one period before moving out. Substituting it in, we obtain

$$\begin{aligned} & \gamma r\left(g(\tau)-g(1)\right)+\beta \int_{X_{\tau+1}}\left[-log\left(1-p(\tau+1, r, Z, X_1, X_{\tau+1})\right)+log\left(1-p(2, r, Z, X_1, X_{\tau+1})\right)\right] dF(X_{\tau+1} \mid X \\ = & log\left(p(\tau, r, Z, X_1, X_1)\right)-log\left(1-p(\tau, r, Z, X_1, X_1)\right)-\log p_r\left(X_1, r, Z, X_1\right) \end{aligned}$$

Since g is known from the rent regulation, and $p()$ and $p_r()$ can be estimated straightforwardly from the data, and if we set β to be a commonly used discount factor, then the only unknown variable in the above equation is γ. Thus, γ is identified. Next, we consider identification of other parts of the model, i.e. u, $b\left(\right)$, $h()$, $c(X)$ and $\eta\left(\right)$. If the observed individual characteristics at tenure τ is X, then

$$\begin{aligned} & \left[b\left(Z, X, u_Z\right) + u_v\left(r - h(Z)\right)\right] c(X) u_c + \gamma r_0 g(\tau) - \eta\left(X, u\right) \\ & + \beta \int_{X_{\tau+1}} \left[\eta\left(X_{\tau+1}, u\right) - \log\left(1 - p\left(\tau+1, r, Z, X_1, X_{\tau+1}\right)\right)\right] dF\left(X_{\tau+1} | X\right) \\ = \; & \log\left(p\left(\tau, r, Z, X_1, X\right)\right) - \log\left(1 - p\left(\tau, r, Z, X_1, X\right)\right) \end{aligned}$$

where

$$\begin{aligned} \eta\left(X, u\right) = \; & \left[b\left(Z\left(X, u\right), X, u_Z\right) + u_v\left(r - h(Z\left(X, u\right))\right)\right] c(X) u_c + \gamma r_0 g(1) \\ & + \beta \int_{X_{\tau+1}} \left[\eta\left(X_{\tau+1}, u\right) - \log\left(1 - p\left(2, r, Z\left(X, u\right), X, X_{\tau+1}\right)\right)\right] dF\left(X_{\tau+1} | X\right) \\ & - \log p_r\left(X, u\right) \end{aligned}$$

Therefore, substituting them in, we obtain as before,

$$
\begin{aligned}
& log\left(p(\tau,r,Z,X_1,X)\right)-log\left(1-p(\tau,r,Z,X_1,X)\right) \\
= \; & \left[b\left(Z,X,u_Z\right)-b\left(Z\left(X,u\right),X,u_Z\right)+u_v\left(r-r\left(X,u\right)-h(Z)+h(Z\left(X,u\right))\right)\right]c(X)u_c \\
& +\gamma\left(r-r\left(X,u\right)\right)g(\tau)+\gamma r\left(X,u\right)\left(g(\tau)-g(1)\right) \\
& -\beta\int_{X_{\tau+1}}\left[\log\left(1-p\left(\tau+1,r,Z,X_1,X_{\tau+1}\right)\right)\right]dF\left(X_{\tau+1}|X\right) \\
& +\beta\int_{X_{\tau+1}}\left[\log\left(1-p\left(2,r\left(X,u\right),Z\left(X,u\right),X,X_{\tau+1}\right)\right)\right]dF\left(X_{\tau+1}|X\right) \qquad (9) \\
= \; & \left[b\left(Z,X,u_Z\right)-b\left(Z\left(X,u\right),X,u_Z\right)+u_v\left(r-r\left(X,u\right)-h(Z)+h(Z\left(X,u\right))\right)\right]c(X)u_c \qquad (10) \\
& +\gamma\left(r-r\left(X,u\right)\right)g(\tau) \qquad (11) \\
& -\beta\int_{X_{\tau+1}}\left[log\left(1-p(2,r,Z,X_1,X_{\tau+1})\right)-\log\left(1-p\left(2,r\left(X,u\right),Z\left(X,u\right),X,X_{\tau+1}\right)\right)\right]dF\left(X_{\tau+1}|X\right) \qquad (12) \\
& -log\left(p(\tau,r,Z,X_1,X_1)\right)+log\left(1-p(\tau,r,Z,X_1,X_1)\right)+\log p_r\left(X_1,r,Z,X_1\right) \qquad (13)
\end{aligned}
$$

Consider different functions $\widetilde{b}\,()$, $\widetilde{h}(Z)$, $\widetilde{c}(X)$ and $\widetilde{\eta}\,()$. Then, using equations 5 and 6, we derive for the same Z, X_1,

$$(\widetilde{u}_Z, \widetilde{u}_v, \widetilde{u}_c)$$

where (u_z, u_v, u_c) is the true individual characteristic. Then, for the individuals who we have information on apartment where they were living and the apartment where they move into, we know that both spells have the same unobserved individual characteristics.

$$
\begin{aligned}
& \left[b\left(Z,X,u_Z\right)-b\left(Z\left(X,u\right),X,u_Z\right)+u_v\left(r-r\left(X,u\right)+h(Z)-h(Z\left(X,u\right))\right)\right]c(X_\tau)u_c \\
& +\gamma\left(r-r\left(X,u\right)\right)g(\tau) \\
= \ & \beta\int_{X_{\tau+1}}\left[log\left(1-p(2,r,Z,X_1,X_{\tau+1})\right)-\log\left(1-p\left(2,r\left(X,u\right),Z\left(X,u\right),X,X_{\tau+1}\right)\right)\right]dF\left(X_{\tau+1}|X\right) \\
& -\left[log\left(p(\tau,r,Z,X_1,X)\right)-log\left(1-p(\tau,r,Z,X_1,X)\right)\right] \\
& +\left[log\left(p(\tau,r\left(X,u\right),Z\left(X,u\right),X,X)\right)+log\left(1-p(\tau,r\left(X,u\right),Z\left(X,u\right),X,X)\right)\right] && (14) \\
& -\log p_r\left(X_1,r,Z,X_1\right) && (15)
\end{aligned}
$$

Similarly as before, we derive

$$
\begin{aligned}
& \beta\int_{X_{\tau+1}}\left[log\left(1-p(2,r,Z,X_1,X_{\tau+1})\right)-\log\left(1-p\left(2,r\left(X,u\right),Z\left(X,u\right),X,X_{\tau+1}\right)\right)\right]dF\left(X_{\tau+1}|X\right) \\
& -\left[log\left(p(\tau,r,Z,X_1,X)\right)-log\left(1-p(\tau,r,Z,X_1,X)\right)\right] \\
& +\left[log\left(p(\tau,r\left(X,u\right),Z\left(X,u\right),X,X)\right)+log\left(1-p(\tau,r\left(X,u\right),Z\left(X,u\right),X,X)\right)\right] \\
& -\log p_r\left(X_1,r,Z,X_1\right) && (16) \\
= \ & \left[b\left(Z,X,u_Z\right)-b\left(Z\left(X,u\right),X,u_Z\right)+u_v\left(r-r\left(X,u\right)+h(Z)-h(Z\left(X,u\right))\right)\right]c(X_\tau)u_c \\
& +\gamma\left(r-r\left(X,u\right)\right)g(\tau) \\
= \ & \left[\widetilde{b}\left(Z,X,\widetilde{u}_Z\right)-\widetilde{b}\left(Z\left(X,u\right),X,\widetilde{u}_Z\right)+\widetilde{u}_v\left(r-r\left(X,u\right)+\widetilde{h}(Z)-\widetilde{h}(Z\left(X,u\right))\right)\right]\widetilde{c}(X_\tau)\widetilde{u}_c \\
& +\gamma\left(r-r\left(X,u\right)\right)g(\tau) && (17)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& \left[b\left(Z, X, u_Z\right)-b\left(Z\left(X, u\right), X, u_Z\right)+u_v\left(r-r\left(X, u\right)-h(Z)+h(Z\left(X, u\right))\right)\right] c(X) u_c \\
= & \left[\widetilde{b}\left(Z, X, \widetilde{u}_Z\right)-\widetilde{b}\left(Z\left(X, u\right), X, \widetilde{u}_Z\right)+\widetilde{u}_v\left(r-r\left(X, u\right)-\widetilde{h}(Z)+\widetilde{h}(Z\left(X, u\right))\right)\right] \widetilde{c}(X) \widetilde{u}_c
\end{aligned}
$$

Now, consider the case where X_1 changes, but u and X remains the same. Then, $z(X,u)$ and $r(X,u)$ also remains the same, but (Z,r) change to (Z',r'). Then, we have

$$
\begin{aligned}
& \left[b\left(Z, X, u_Z\right)-b\left(Z', X, u_Z\right)+u_v\left(r-r'-h(Z)+h(Z')\right)\right] c(X) u_c \\
= & \left[\widetilde{b}\left(Z, X, \widetilde{u}_Z\right)-\widetilde{b}\left(Z', X, \widetilde{u}_Z\right)+\widetilde{u}_v\left(r-r'-\widetilde{h}(Z)+\widetilde{h}(Z')\right)\right] \widetilde{c}(X) \widetilde{u}_c
\end{aligned}
$$

By comparing the coefficients of $r-r'$, we obtain

$$
c(X) u_c u_v=\widetilde{c}(X) \widetilde{u}_c \widetilde{u}_v
$$

Because $u_c u_v$ and$\widetilde{u}_c \widetilde{u}_v$ do not vary with X, for a positive constant A

$$
\widetilde{u}_c \widetilde{u}_v=A u_c u_v, \;\; \widetilde{c}(X)=\frac{c\left(X\right)}{A}
$$

Furthermore,

$$
\begin{aligned}
& b\left(Z, X, u_Z\right)-b\left(Z', X, u_Z\right)+u_v\left(-h(Z)+h(Z')\right) \\
= & \left[\widetilde{b}\left(Z, X, \widetilde{u}_Z\right)-\widetilde{b}\left(Z', X, \widetilde{u}_Z\right)\right] \frac{u_v}{\widetilde{u}_v}+u_v\left(-\widetilde{h}(Z)+\widetilde{h}(Z')\right)
\end{aligned}
$$

Now, we know that

$$
\begin{aligned}
u_c u_v &= \frac{h_{Z_j}(Z)\left[\gamma_1+\gamma_2 r(1)+\sum_{\tau=2}^{T}\beta^{\tau-1}\int_{X_{1:\tau}} p\left(s>\tau, r\left(0\right), Z, X_1, X_\tau\right) dF\left(X_\tau|X_1\right)\left[\gamma_1+\gamma_2 r(0) g(\tau)\right] g(\tau)\right]}{c(X_1)+\sum_{\tau=2}^{T}\beta^{\tau-1}\int_{X_{1:\tau}} p\left(s>\tau, r\left(0\right), Z, X_1, X_\tau\right) c(X_\tau) dF\left(X_\tau|X_1\right)} \\
&= \frac{h_{Z_j}(Z)\left[\gamma_1+\gamma_2 r(1)+\sum_{\tau=2}^{T}\beta^{\tau-1}\int_{X_{1:\tau}} p\left(s>\tau, r\left(0\right), Z, X_1, X_\tau\right) dF\left(X_\tau|X_1\right)\left[\gamma_1+\gamma_2 r(0) g(\tau)\right] g(\tau)\right]}{A\left[\widetilde{c}(X_1)+\sum_{\tau=2}^{T}\beta^{\tau-1}\int_{X_{1:\tau}} p\left(s>\tau, r\left(0\right), Z, X_1, X_\tau\right) \widetilde{c}(X_\tau) dF\left(X_\tau|X_1\right)\right]} \\
&= \frac{\widetilde{u}_c\widetilde{u}_v}{A} = \frac{\widetilde{h}_{Z_j}(Z)\Big[\gamma_1+\gamma_2 r(1)+\sum_{\tau=2}^{T}\beta^{\tau-1}\int_{X_{1:\tau}} p\left(s>\tau, r\left(0\right), Z, X_1, X_\tau\right) dF\left(X_\tau|X_1\right)\big[\gamma_1+\gamma_2 r(0) g(\tau}{A\left[\widetilde{c}(X_1)+\sum_{\tau=2}^{T}\beta^{\tau-1}\int_{X_{1:\tau}} p\left(s>\tau, r\left(0\right), Z, X_1, X_\tau\right) \widetilde{c}(X_\tau) dF\left(X_\tau|X_1\right)\right]}
\end{aligned}
$$

Because this holds for any r, Z given $u_c u_v$ and $\widetilde{u}_c\widetilde{u}_v$,

$$h_{Z_j}(Z) = \widetilde{h}_{Z_j}(Z)$$

Hence,

$$b\left(Z, X, u_Z\right) = \widetilde{b}\left(Z, X, \widetilde{u}_Z\right)\frac{u_v}{\widetilde{u}_v}$$

The parameters are estimated on the mobility data. That is, parameters are chosen so that the generated unobserved heterogeneity results in per period utility that explains best the pattern of mobility. That is, parameters should be chosen such that individuals who stayed longer in an apartment have higher per period utility. If we can assume that the unobserved characteristics v_i are orthogonal to the observed ones, Z_i, then, similarly to the standard hedonic literature, we can identify most of the parameters of the hedonic model without duration data, using the static hedonic model. To see this, first consider the rent equation. With the additional orthogonality condition, we can obtain the function $h(Z)$ of the rent equation just by OLS. Then, if we modify equation (4) and add another term $b_x X_i$, it becomes

$$r_v b_Z(Z_i, X_i) u_Z + r_v b_x X_i = h'(Z_i) \tag{18}$$

This setup is very similar to **?** and **?** where the authors show identification of the preference parameters given that the utility shock u_Z is independent of X_i, and given some functional form assumptions. It is of some interest that in static hedonic model, one can only identify the utility function parameters well if the marginal utility function is fully flexible. One common restriction is that it is linearly separable in Z_i and X_i. However, we have shown that in dynamic hedonic models, the utility function component $b(Z_i, X_{i,\tau})$ can be made more flexible. In sum, both the hedonic optimal choice part and the dynamic discrete mobility choice part complement each other in identifying the structural parameters and the unobserved heterogeneity of each individual.

4 Estimation

The model is estimated by a two step approach maximum likelihood approach. We present in the next section of estimation algorithm, then we proceed with the functional form assumption. It is useful to provide the intuition. Individual decision in our model is based on a dynamic programming that is computationally intensive. Our estimation proposed a method designed to overcome this issue in several informal steps. It is clear from equation (2), that the probability to move out can be easily recovered. Since, we assume optimality in the housing choice of individuals, equations (**??**) and (5) allow to recover the shocks using only the moving-out probability. The final issue is related to the value of moving. Since, we do not know the destination of the individuals, we use the value of staying at initial period of individuals with the same observed characteristics.

4.1 Estimation algorithm

The presence of utility shocks in the value of moving creates a nontrivial estimation problem. Our approach takes the most of the data in order to recover all the parameters consistently. The first step is based on the idea that individuals choose both observed and unobserved characteristics of the housing unit in the initial period. As a consequence, we can write

$$Z = Z(x_0, u_Z, u_i)$$

Then from inversion, an expression for the housing and individual unobserved u_Z and u_i can be obtained:

$$u_Z, u_i = u(X_0, Z)$$

Therefore, the survival probability can be expressed as

$$P(s \geq t, u_Z, u_i, \{X_{i,\tau}\}_{\tau=1}^{T}, Z) = P(s > t, \{X_{i,\tau}\}_{\tau=1}^{T}, Z) \tag{19}$$

Given the survival probability, the first step choice and using equations (5) and (**??**), we can write a close form expression for u_{vi} and u_i.

$$u_v = \frac{\sum_{l=1}^{T} \beta^l E\big[P(s > l, \{X_{i,l}\}_{l=1}^{T}) b_{Z_J}(Z_i, X_{i\tau}, u_Z) c(X_{i,l}) \mid X_{i0}\big]}{\sum_{l=1}^{T} \beta^l E\big[P(s > l, \{X_{i,l}\}_{l=1}^{T}) c(X_{i,l} \mid X_{i0})\big] h_{Z_J}(Z_i)} \tag{20}$$

$$u_i = -\frac{h_{Z_J}(Z) \sum_{l=1}^{T} \beta^l E\big[P(s > l, \{X_{i,l}\}_{l=1}^{T}) \mid X_{i0})\big][\, \gamma_1 + \gamma_2 r_i(0) g(l)] g(l)}{\sum_{l=1}^{T} \beta^l E\big[P(s > l, \{X_{i,l}\}_{l=1}^{T}) b_{Z_J}(Z_i, X_{il}, u_Z) c(X_{i,l}) \mid X_{i0}\big]} \tag{21}$$

The second step uses the former shocks for estimating the structural parameters of the model.

We use the following formula:

$$EV_\tau = \log\big[\exp(\bar{V}_{s\tau}) + + \exp(\bar{V}_{m\tau})\big]$$

Since

$$1 - p_m = \frac{\exp\big[\bar{V}_s\big]}{\exp[\bar{V}_s] + \exp[\bar{V}_m]} \tag{22}$$

We can derive

$$EV_\tau = \bar{V}_s + \log\big[1 - p_m\big]$$

Then, from the dynamic programming,

$$\bar{V}_{s,\tau} = \gamma_1 r(\tau, Z_i, v_i) + \frac{1}{2}\gamma_2 r(\tau, Z_i, v_i)^2 + \zeta_{i0} + + \beta EV_{\tau+1} \tag{23}$$

By backward induction, $\{\bar{V}_{s\tau}\}_{\tau=1}^{T}$ can be derived. We can then move to the derivation of $\bar{V}_{m\tau}$. From the equation involving the choice probability

$$\bar{V}_{m\tau}(X_{i\tau}, Z_i, u_v, u_Z, u_i) = \bar{V}_{s\tau} + \log(p_{m\tau}) - \log(1 - p_{m\tau}) \tag{24}$$

The idea behind the estimation is the following: when an individual decides to move, she chooses a new apartment. From the choices of similar individuals in the sample (u_i,u_v,u_Z and X) at the mobility time, we get to know

$$\bar{V}_{m\tau}(X_{i\tau}, Z_i, u_v, u_Z, u_i) = \bar{V}_{s0}(X_{i0} = X_{i\tau}, Z_i, u_v, u_Z, u_i)$$

Then, we choose the parameters such that:

$$\bar{V}_{s\mathcal{T}} + \log(p_{m\mathcal{T}}) - \log(1 - p_{m\mathcal{T}}) - \bar{V}_{s0}(X_{i0} = X_{i\mathcal{T}}, Z_i, u_v, u_Z, u_i) = 0$$

More formally, the problem can be rewritten:

$$\bar{V}_{s\mathcal{T}} + \log(p_{m\mathcal{T}}) - \log(1 - p_{m\mathcal{T}}) - \sum_i \bar{V}_{s0}(X_{i0} = X_{i\mathcal{T}}, Z_i, u_v, u_Z, u_i) K(u_{Z,i} - u_Z) K(u_{v,i} - u_v) K(u_{i,i} - u_i) = 0$$

Where $K()$ is a kernel allowing to match individuals.

5 Results

5.1 Baseline results for basic model

5.2 Preferences and Rent parameters

In Table 3, we report the maximum likelihood estimation results of the basic model. All structural parameters have the expected sign. The price coefficients γ_1, for the per period utility of staying is negative and significant. The square price coefficient γ_2 is negative and significant. The coefficient of the squared term of the number of rooms, and mean size per room are positive and significant indicating that higher number of rooms, and bigger rooms yield a higher utility. Moreover, the coefficient of the dummy for two bathrooms is positive and significant but of lower magnitude than the number of rooms. The coefficient that measures the interaction between age and number or rooms in the utility function is negative, implying that marginal utility of number of rooms is increasing with age that is consistent with the life-cycle theory of housing demand. The coefficients of the utility of moving out are both significant. The coefficient of log age is negative, which reflects the data where older people move less often. Finally, the coefficients in the rent equation are one would have expected.

In monetary amount, the model states that one needs 121 € for an additional room, and 111 € for 10 m^2 of additional floor area while a second bathroom costs 303 €. In terms of rent curve, the model implies a price increase of 47 € for additional 100,000 inhabitants while the rent decreases of 21 € for each additional kilometre to the center. These parameters are in contrast with the very simple version of the typical price equation estimated in the hedonic literature. We can see that the OLS estimated coefficients are much lower, from about 20% to 90% of the value of the coefficient estimated by ML. The ML and OLS results imply that the unobserved housing specific heterogeneity is negatively correlated with the observed characteristics included in this estimation.

6 Conclusion

In this paper, we proposed to estimate the parameters of the price equation and the structural hedonic model by utilizing the tenure and mobility data of rental apartments in France. We used the French data on rental apartments, because it has the detailed information on tenure, i.e. length of stay in an apartment and mobility, in addition to the detailed information on the characteristics of the apartment and the renter. The additional benefit of the data on French apartments are that in France, the rent is regulated to grow at a low mandated rate, which is not related to the local housing market. Therefore, we do not need to worry about the endogeneity of the price of rent when we use the over time variation of rent for identification of the price coefficient. The literature has estimated dynamic models of housing choice, and also estimated the duration models on the tenure choice, but so far it has not explicitly used the duration data to identify the parameters of the hedonic model.

It turns out that the coefficients of the observed housing characteristics are identified without the restriction that the observed and unobserved characteristics are orthogonal, the assumption that is often used in estimating the hedonic model. Furthermore, marginal utility of observed characteristics can be a flexible function of observed housing and individual characteristics. This is in contrast to the conventional identification and estimation strategy of hedonic models, where marginal utility needs to be a separable function of observed housing characteristics and the characteristics of the consumer

The estimation results demonstrate that the observed and unobserved housing characteristics are negatively correlated, which is reasonable, since what we in general see in France is that in expensive neighborhoods, renters live in apartments with lower observed quality to make up for the high unobserved qualities of the neighborhood. Failure to take into account the endogeneity bias could underestimate the true value of the observed characteristics of housing.

An interesting direction for future research would be to extend the analysis of hedonic dynamic model to explicitly include neighborhood effects. There has been much interest in estimating the neighborhood effects in the static and dynamic hedonic literature. However, since neighborhood unobserved heterogeneity is also part of unobserved heterogeneity of a housing unit, we believe that the proper identification of it can be only done as an extension of the identification and estimation of the individual unobserved housing characteristics. This extension is left for future research.

References

A Notations

B Data selection

Our data source is the French subset of European Community Statistics on Income and Living Conditions (EU-SILC) on the period from 2004 to 2009. As explained before, the sample is a rotating panel. Table 5 presents the sample size per year and how individuals should exit over time. Table 6 analyzes the mobility patterns of individuals. And finally, Table 7 summarizes the number of years spent by individuals in the survey.

Table 1: Transitions in the Housing Market

		Mobile renters			Non-mobile
		RH	RR	RP	renters
Age	18-24	10.2	23.3	10.5	7.8
	25-29	22.8	20.4	20.9	9.6
	30-34	28.0	18.7	18.6	12.0
	35-39	15.4	10.1	7.0	10.7
	40-44	8.3	7.6	5.8	11.6
	45-49	2.8	5.4	12.8	9.2
	50-54	4.7	3.7	4.7	8.3
	55-59	4.3	2.2	2.3	7.5
	60+	3.5	8.6	17.4	23.4
Male		66.9	60.7	50.0	56.3
Marital Status	Partnership	74.8	45.9	50.0	41.6
	Single	25.2	55.0	50.0	58.4
Family size	1	18.5	39.8	23.3	45.3
	2	38.6	34.4	32.6	27.8
	3	17.7	13.3	24.4	13.8
	4	16.9	7.1	11.6	8.4
	5+	8.3	5.4	8.1	4.7
Number of rooms	1	4.3	16.7	15.3	13.2
	2	21.7	28.7	27.1	25.2
	3	30.3	28.5	30.6	28.9
	4	25.2	13.8	12.9	21.8
	5+	18.5	12.3	14.1	10.9
Floor area in m^2		74.7	64.2	62.6	67.8
Mean income in €		3541.9	2103.3	1959.0	2041.5
Initial rent in €		554.4	470.1	391.8	463.7
Sample size		254	407	86	774

Sources: SILC (2004-2009)
Notes: RH, RR and RP refer respectively to transitions from private renting to homeownership, private renting to renting and private renting to public housing.

Table 2: OLS

Parameters	Parameters	S.E.
Const	1.11***	0.13
Number of rooms	0.99***	0.02
Mean size per room	0.70***	0.04
Dummy for two bathrooms	2.12***	0.11
Price in the center	0.22***	0.06
Distance to the center	-0.03***	0.002

***, ** and * indicate significance at the 1%, 5% et 10% level

Table 3: Parameter Estimates and Standard Errors

	Parameters	Est.	S.E.
Utility	Age	-0.361*	0.22
	Number of rooms	1.824***	0.28
	Mean size per room	0.292*	0.17
	Dummy for two bathrooms	0.351***	0.10
	Price in the center	0.117	0.08
	Distance to the center	-0.678***	0.04
	γ_1	-0.268***	0.11
	γ_2	-0.001	0.002
Interactions	Age $\times$ # of rooms	-2.993*	1.56
	Age $\times$ Mean size per room	-0.442	0.96
	Age $\times$ Dummy for two bathrooms	0.276	0.34
	Age $\times$ Price in the center	-0.303	0.22
	Age $\times$ Distance to the center	0.440***	0.16
Rent	Number of rooms	1.208***	0.47
	Mean size per room	1.118**	0.51
	Dummy for two bathrooms	3.032***	0.36
	Price in the center	0.473***	0.21
	Distance to the center	-0.217*	0.11
	φ_0	-1.935***	0.51
	φ_1	-0.209***	0.07

***, ** et * indicate significance at the 1%, 5% et 10% level

Table 4: List of notations

Variables	Definition
$b()$	Function of housing unobserved and unobserved that affects the utility
$C()$	Function of individual observed attributes
$h()$	Function of housing unobserved and unobserved that affects the rent
$MC()$	Mobility cost function
X	Individual characteristics
Z	Housing observed characteristics
w_i	Housing unobserved characteristics
ζ	Per period utility of a house
v_i	Utility component of housing unobserved
u_v	Proportionality factor between w_i and v_i
u_Z	Shock to housing observed characteristics
u_w	Shock to housing unobserved characteristics
u_i	Individual shock
τ	Duration
β	Discount factor
γ_1	Rent coefficient in the utility function
γ_2	Square rent coefficient in the utility function
V_m	Value of moving
V_s	Value of staying

Table 5: Attrition and Exit in the Sample

		Year of arrival in the survey					
		2004	2005	2006	2007	2008	2009
Observed Attrition year	2005	1320	-	-	-	-	-
	2006	1675	324	-	-	-	-
	2007	1213	161	292	-	-	-
	2008	1608	143	129	238	-	-
	2009	1060	262	133	135	200	-
Planned Attrition year	2005	1188	-	-	-	-	-
	2006	1162	-	-	-	-	-
	2007	1206	-	-	-	-	-
	2008	1243	-	-	-	-	-
	2009	1199	-	-	-	-	-
Sample size		9,643	1,965	1,959	1,899	1,929	1,852

Table 6: Mobility in the sample

		Year of arrival in the survey					
		2004	2005	2006	2007	2008	2009
Year of mobility	2005	520	0	0	0	0	0
	2006	319	94	-	-	-	-
	2007	207	75	111	-	-	-
	2008	122	49	81	89	-	-
	2009	55	37	40	88	104	-

Table 7: Number of years in the survey

		Homeowners	Private Renting	Public Housing
Years in the survey	1	1255	401	286
	2	2061	817	595
	3	2068	636	528
	4	1769	475	448
	5	1927	419	430
	6	2069	551	502
Sample size		11,149	3,299	2,789

Our final data selection is as follows:

- We dispose of individuals who were homeowners or renter in public housing at the initial period.
- We get rid of individuals who were older than 70 and those with a tenure length higher than 30 years.
- We dispose of individuals with missing length of stay, housing occupation status.
- We do not consider individuals who stayed only one year in the panel.